Capstone data science project

Aug 2020

The problem

- The health service in Seattle City want to understand the amount of emergency vehicles and staff they should hire to deal with road traffic accidents.
    - If they can use a model to predict the severity of road traffic accidents then they can optimise the number of support vehicles and staff they have to deal with the accidents.
    - This can help the health services of Seattle City optimise their resources and have a better chance of saving lives.

The idea

- Using Machine Learning techniques we will build a series of classification models in order to predict the severity of an accident.

- We will test a series of logistic regression and decision trees models in order to understand which model predicts the severity of an accident most clearly.

- Understanding the factors which predict the severity of accidents can help the health services of Seattle City plan for accidents more effectively.

- All models will be built in python using data from Seattle City Health service between 2004 and 2020

    - Seattle city accident data: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

The Models

- We used two classification techniques for this project: decisions trees and logistic regression.
- Both techniques are popular machine learning algorithms which build a model from the historical data of accidents and then predict the severity class of an unknown accident given information.
- These techniques are required in this project as we are using a nominal variable (i.e. predicting if an accident would either be property damage or personal injury). Linear regression cannot be used as we are not predicting a continuous variable.

The Result

- Accident severity can be predicted with an accuracy of 74% using the following variables:
    - Weather
    - Daylight Hours
    - If the vehicle was speeding at the time of the accident
    - If a bicycle was involved in the accident
    - If the road was wet, icy or dry
    - If there was fog when the accident happened
    - The type of accident and the number and vehicles people involved

- The Decision tree and logistic regression gave similar results but the statistical tests for the decision tree were more robust
    - The Decision tree is the best classifier

The Conclusion

- Using accident severity data we have built a robust classification model to help the health service of Seattle predict the severity of an accident given certain conditions such as wet and icy weather.

- This information can help the health service plan their resources and respond more effectively to accidents.

- Modelling is an iterative process and there may be other information which can help to improve the accuracy of the model further.